



Li, G. and Khan, M. A. (2019) Deep Learning on VR-Induced Attention. In: 2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), San Diego, CA, USA, 9-11 Dec 2019, pp. 163-1633. ISBN 9781728156040 (doi:[10.1109/AIVR46125.2019.00033](https://doi.org/10.1109/AIVR46125.2019.00033))

There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

<http://eprints.gla.ac.uk/224102/>

Deposited on 26 October 2020

Enlighten – Research publications by members of the University of Glasgow  
<http://eprints.gla.ac.uk>

# Deep Learning on VR-induced Attention

Gang Li\*, Muhammad Adeel Khan

Department of Micro/Nano Electronics, Shanghai Jiao Tong University  
Shanghai, China

e-mail: lixiaogang110217@hotmail.com

**Abstract**—Some evidence suggests that virtual reality (VR) approaches may lead to a greater attentional focus than experiencing the same scenarios presented on computer monitors. The aim of this study is to differentiate attention levels captured during a perceptual discrimination task presented on two different viewing platforms, standard personal computer (PC) monitor and head-mounted-display (HMD)-VR, using a well-described electroencephalography (EEG)-based measure (parietal P3b latency) and deep learning-based measure (that is EEG features extracted by a compact convolutional neural network—EEGNet and visualized by a gradient-based relevance attribution method—DeepLIFT). Twenty healthy young adults participated in this perceptual discrimination task in which according to a spatial cue they were required to discriminate either a “Target” or “Distractor” stimuli on the screen of viewing platforms. Experimental results show that the EEGNet-based classification accuracies are highly correlated with the  $p$  values of statistical analysis of P3b. Also, the visualized EEG features are neurophysiologically interpretable. This study provides the first visualized deep learning-based EEG features captured during an HMD-VR-based attentional task.

**Index Terms**— Attention, Head-mounted Virtual Reality, EEG, EEGNet, DeepLIFT

## I. INTRODUCTION

ATTENTION is a fundamental cognitive process that is critical for essentially all aspects of higher-order cognition (such as working memory) and real-world activities (such as academic performance) [1]. Decades of research have shown that selectivity, which involves the abilities of focusing on relevant and ignoring irrelevant information, is attention’s most fundamental feature [2]. More importantly, some evidence suggests that focusing and ignoring are not two sides of the same coin; they are two separate coins (That is they are using two different brain networks) [3].

In this context, head-mounted-display (HMD)-VR elevates the cognitive science research on attention to the next level. This is because: 1) HMD-VR’s uniqueness of completely blocking out the physical world improves the ability of ignoring irrelevant information naturally if compared to a standard personal computer (PC) screen; 2) The high degree of immersion that HMD-VR offers results in strong sense of presence by increasing a user’s allocation of attentional resources [4]. Therefore, HMD-VR provides neuroscientists an unprecedented understanding of how attention abilities could be maximized in the context of fully immersive virtual environment. Actually, there has been some neural evidence to

support these VR-related arguments. Participants performing a navigation task in a semi-immersive VR environment exhibited enhanced spectral features of EEG (FFT-based midline frontal theta power) [5] and time-locking features (frontal slow wave component of event-related potential (ERP) [6]) if compared to non-immersive VR. However, other researchers have found a decreased trend in phase-locking features of EEG (frontal-parietal coherence across the theta and alpha frequency bands) [7], suggesting that these neural relationships may not be quite so clear. It is likely that differences in the quality of attentional tasks themselves contribute to these inconsistencies across studies. For example, none of them used the proven paradigms (such as Posner task) to assess the VR-induced attention.

Recently, deep learning has been applied to analyze EEG data [8]. With its advantage of automatically learning features from EEG data, deep learning-based approaches have achieved comparable accuracy in emotion recognition [8] to those handcrafted features. Such great breakthrough is an important step towards making the use of EEG more practical in many applications and less reliant on trained professionals. However, the lack of feature visualization and explainability makes us hard to tell that these encouraging results are achieved by real interpretable neural features or noise and artifacts contained in the data.

The goal of this study was to differentiate attention levels captured during a proven perceptual discrimination task presented on two different viewing platforms, standard personal computer (PC) monitor and HMD-VR, using a well-described ERP measure (parietal P3b latency) and deep learning-based measure (that is EEG features extracted by EEGNet [9] and visualized by DeepLIFT [10]). Given the higher degree of immersion that HMD-VR is expected to offer (compared to the standard PC screen), we hypothesized that participants engaged in the HMD-VR task would generate enhanced neural evidence of attention (such as shorter P3b latency), associated with those features extracted by EEGNet.

## II. METHODS AND MATERIALS

### A. Perceptual Discrimination Task

As shown in Fig. 1, we developed an HMD-VR game, delivered in HTC Vive™—a flagship consumer-friendly HMD-VR platform powered by graphic card of NVIDIA GeForce GTX 1070, to assess the selective attention abilities in the form of a perceptual discrimination task, with the ability to collect simultaneous EEG recordings. This game is developed from the principles of a previous cognitive assessment/intervention (DAT [11]), which was based upon a traditional Posner task [12]. To evaluate the hypothesis that the HMD-VR would be more attention engaging, we

This research is sponsored by Shanghai Sailing Program under Grant 17YF1426900.

\*corresponding author: G. Li; e-mail: lixiaogang110217@hotmail.com.

## Proposed VR-EEG game platform

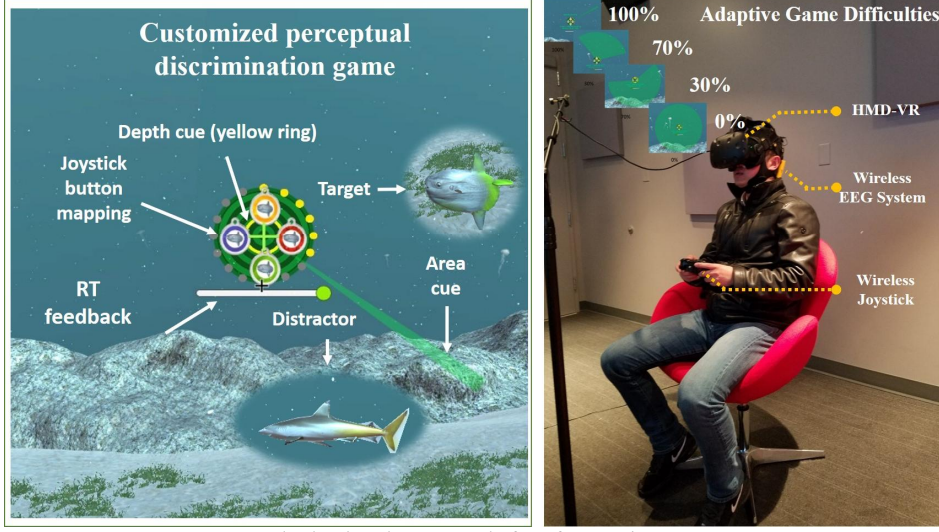


Fig. 1. The developed VR-EEG platform for attention assessment.

compared performance on this device with a standard PC 2D monitor (fixed viewpoint).

Each trial of the game begins with the appearance of a composite cue indicating the area (parallel to PC screen/HMD screen) and the depth (vertical to PC screen/HMD screen) where a single ocean animal (either a “Target” or “Distractor” stimuli) would appear. Both Target and Distractor stimuli were presented in a pseudo-randomized fashion on a trial-by-trial basis. The composite cue consisting of area (a light green sector) and depth information (a yellow ring) appeared on the screen for 300 msec. Upon the appearance of these stimuli, participants were instructed to press a button if it was a target (a sunfish) and to not release their thumb from the home position if it was a distractor (any other fish). As the increase in game difficulty, the diameter of the yellow ring will become larger and the light green sector will finally become a 360° circle, indicating less and less spatial information of the stimuli. Thus, the harder to predict where the stimuli would appear. The interval between cue and ocean animal presentation was set at 1.2 sec. The inter-trial interval was set at 1.5 sec.

We used a 20-channel wireless EEG recording device (Enobio™), which uses a high-resolution analog-to-digital converter (24bit at 500Hz sampling rate), and supports WiFi connection. The conventional wet electrodes were used and placed at all 20 channels including frontal (Fp1, Fp2, Fz, F3, F4), central (C3, Cz, C4), temporal (T7, T8), parietal (P3, P4, P7, Pz, P8) and occipital (O1, O2) regions. The ground and reference electrode were connected together and placed on the right earlobe by an ear clip. An external electrode (EXT) was placed below the lower eyelid to record eye movements.

### B. EEG Data Pre-Processing

A low-pass filter with a cutoff frequency of 30 Hz and high-pass filter with a cutoff frequency of 0.5 Hz were applied to remove power line noise and the DC drift, respectively. The filtered EEG signals were then corrected using the mean of

each channel. The prominent artefactual components, such as eye blinks, eye movements and muscle activity were removed. Next, the target/distractor epochs of -200ms to 600ms were created and further cleaned using a voltage threshold of 100uV.

### C. P3b

All ERPs were baseline-corrected using a -200ms to 0ms time period, with the window of interest interrogated being 250-500ms post-stimulus for P3b—a ERP component which is hypothesized to reflect allocation of attention resources [13], and has been shown highly correlated with motor action, such as pressing a button [4]. Thus, we used P3b as our traditional neural marker to evaluate the attention level. Given our focus on response time-based metrics, we focused on P3b latency in the Pz channel, which is the location that the P3b is commonly reported to reach its maximum amplitude [1].

### D. EEGNet and DeepLIFT

EEGNet is a lightweight 2D convolutional neural network (CNN), which was used here to extract features from pre-processed EEG time series and further classify these extracted features into two classes (That is 2D/VR-induced attention) or four classes (That is the same trial type between platforms: 2D-Target, VR-Target, 2D-Distractor, and VR-Distractor). The EEGNet architecture can be found in Table II in [9]. The details of the input to EEGNet and the training parameters are described below. Note that before these detailed information and parameters were determined we compared the EEGNet classification accuracy with another lightweight 2D CNN claimed in [15]. (See *Results* part).

#### 1) Input to the Network

The input to EEGNet was the tensor ( $N \times 1 \times C \times T$ ), where  $N$  denotes the total number of 800-ms long EEG segments obtained in *Pre-Processing* part ( $N=6141$ );  $C$  and  $T$  denote the number of channels ( $C=19$ , EXT excluded) and time samples ( $T=103$ , in the context of down sampling ( $F_s$ ) to 128Hz). All

EEG samples have been normalized using z-score before segmented. All class labels have been converted to binary class metrics using hot encoding. Before training, those EEG segments were randomized to either training dataset or test dataset according to a ratio of 80/20. Therefore, a total of 4606 EEG segments were used for training, and 1535 for final test.

### 2) Training

The training of EEGNet was based on the Adam optimizer, aimed at minimizing the binary (for two-class) or categorical (for four-class) cross-entropy loss function. The activation function of dense layer was *sigmoid* for two-class and *softmax* for four-class classification. We run 500 training iterations and 4-fold cross-validation. All models were trained and tested on the same GPU for HMD-VR, with CUDA 10 and cuDNN v10, in Anaconda-powered Tensorflow and Keras API. The dropout and batch size were set to be 0.5 and 128, respectively. Other parameters are summarized in Table I.

TABLE I  
SUMMARY OF THE TRAINING PARAMETERS FOR EEGNET

Standard terminology	EEGNet's terminology	Values
Number of Temporal filters	$F_1$	128
Length of the Kernel	$kernLength$	$F_2/2$
Depth Multiplier	$D$	2
Number of Pointwise filters	$F_2$	$F_1 \times 2$

### 3) Visualizing

To visualize the EEGNet-based features, DeepLIFT, a gradient-based relevance attribution method that calculates relevance values per feature on the resulting classification decision, was used in this study. Positive values of relevance denote evidence supporting the outcome, while negative values of relevance denote evidence against the outcome.

### E. Study Design, Statistical Analysis and Participants

The experimental design was a within-group randomized approach, with participants completing 4 VR runs and 4 2D runs. Each run contained 50 "Target/Distractor" trials, equally divided between Target and Distractor trial types presented randomly with no more than 4 consecutive trial type of either kind in a row. These parameters were used in previous studies where a perceptual discrimination task was utilized to assess attention-based processes [11]. Between each run, participants were given a 2-min break time. All P3b data were analyzed using standard two-way repeated ANOVA with platform (VR/2D) and trial type (Target/Distractor) as within-subject factors. Paired t-tests were used to further compare the participant's performance on the same trial type between platforms. The correlates of  $p$  value of P3b and EEGNet-based accuracy were investigated using Pearson coefficient. A total of 46 interested university students signed up for this study through our online advertisements. With power analysis (GPower v3.1), we calculated that  $n=16$  would yield 95% power to detect a change with a medium effect size (0.5). Thus, a reasonable sample size ( $n=20$ , healthy 20-25 years old, 4 females) were further invited to schedule a lab session. All participants were paid \$15/hr for their participation and gave

written informed consent before participation.

## III. RESULTS

### A. P3b

We did observe significant main effect of platform for P3b latency ( $F(1, 19)=17.003$ ,  $p=0.001$ ,  $\eta^2=0.472$ ). Pairwise comparisons of the main effect revealed that the latencies in VR platform were shorter than those generated in the 2D platform. Furthermore, as shown in Fig 2, the latency under VR-Target condition was significantly lesser ( $t(19)=-3.337$ ,  $p=0.003$ ) than the 2D-Target condition, with the same pattern for the Distractor trials ( $t(19)=-2.779$ ,  $p=0.012$ ). These results suggest that participants' posterior area in the VR platform had improved attention than those in 2D environment.

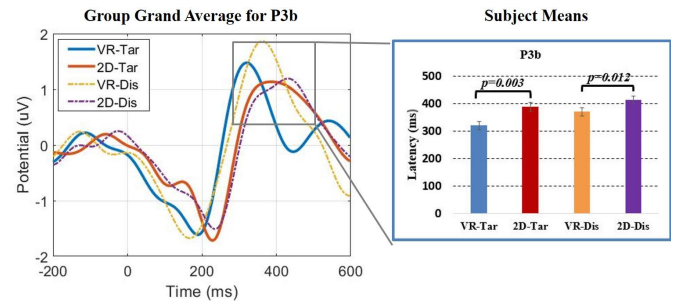


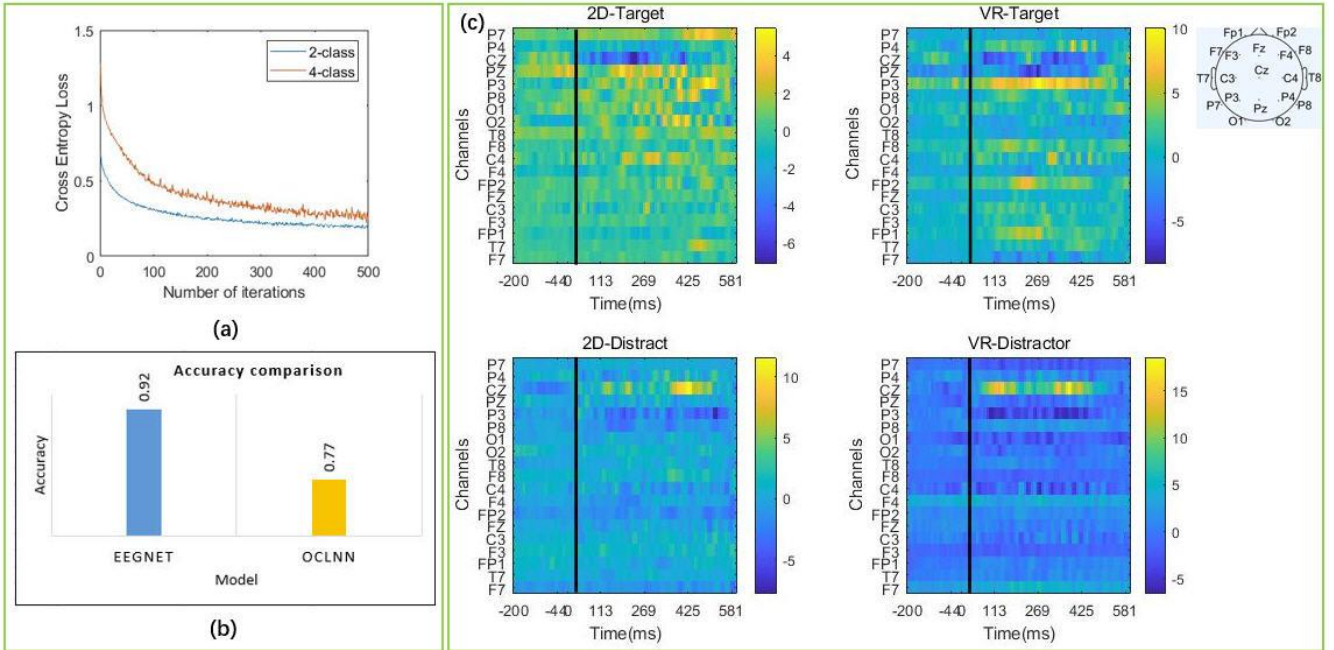
Fig. 2. The group grand average for P3b and corresponding subject means

### B. EEGNet-based Classification and Features

EEGNet-based approach achieved a best 91.86% 4-fold cross-validation accuracy for 2D and HMD-VR-induced two-class classification of attention level, associated with significant difference (ANOVA,  $p=0.001$ ) in main effect of platform for P3b. Learning process of the EEGNet model is shown in Fig. 3(a) in terms of cross entropy loss during training phase for 500 iterations across 6141 EEG segments from 20 participants. Apparently, the model fits the data at highest rate in the first 100 iterations, then the validation loss is approaching steady state at around the 500<sup>th</sup> iteration. The number of iterations is the same as its default value in [9]. This result is superior to another lightweight CNN model proposed in [16] which achieved only 77% in this study (as shown in Fig. 3(b)). Also, EEGNet achieved a best 85.52% accuracy for four-class classification of attention level (83% for 2D-Target, 83% for 2D-Distractor, 91% for VR-Target, and 86% for VR-Distractor), associated with significant difference in VR-Target versus 2D-Target (paired t-test,  $p=0.003$ ) and VR-Distractor versus 2D-Distractor (paired t-test,  $p=0.012$ ) for P3b. The Pearson coefficient between  $p$  values of P3b and EEGNet-based accuracies is -0.83, indicating the noticeable links between deep learning-based approach and the traditional statistical analysis.

Understanding how EEGNet achieves such performance is of equal importance. First, as shown in Fig. 3(c), all peaks of relevance values are appeared at the post-stimulus timing, indicating the classification results may indeed be driven by real EEG features rather than noise and artifacts. Second, we found that more frontal brain regions engaged in Target trial type for both VR and 2D platform if compared to Distractor trial type, indicating greater attentional resource when





**Fig. 3. (a)** Learning curve of two-class (2D/VR) and four-class classification (the same trial type between platforms) in 500 training iterations. **(b)** Two-class classification (2D/VR) accuracy comparison between EEGNet and OCLNN proposed in [16]. **(c)** The EEG features induced by the same trial type under 2D/VR platform and then extracted by EEGNet and further visualized by DeepLIFT.

attending, and lesser when ignoring. This has been a well-known conclusion in neuroscience field [2]. Third, apparently, the peak of the relevance values for VR platform (colorbar: 10 for VR-Target and 15 for VR-Distractor) are higher than that in 2D platform (colorbar: 4 for 2D-Target and 10 for 2D-Distractor), indicating the VR-induced heightened neural evidence if compared to 2D. Fourth, for Target trial type, the high relevance values in VR platform are more channel-focused and long-lasting as opposed to those that are scattered and intermittent in the 2D platform, indicating that HMD-VR may indeed be better than 2D platform when directing our attention in a focused manner. All these results show that EEGNet-based features are neurophysiologically interpretable.

#### IV. CONCLUSION

The present findings reveal that the attention level is improved in young adults when a gamified perceptual discrimination task is executed using HMD-VR platform (if compared to 2D platform), evidenced via both traditional EEG features and deep learning-based features. To our knowledge, this study provides the first visualized deep learning-based EEG features captured during a HMD-VR-based selective attention paradigm task. Future research will be needed to explore how well the HMD-VR-related attentional benefits are replicated by larger and more diverse populations by using the burgeoning all-in-one HMD-VR platform (e.g., Oculus Quest).

#### References

- [1] D. A. Ziegler, et al, "Closed-loop digital meditation improves sustained attention in young adults," *Nature Human Behavior*, Vol. 3, No. 6, June 2019.
- [2] A. Gazzaley, and L. D. Rosen, "The Brain and Control," in *The Distracted Mind*, Cambridge, MA: MIT Press, pp.54-59, 2016.
- [3] J. Z. Chadick, T. P. Zanto, and A. Gazzaley, "Structural and Functional Differences in Medial Prefrontal Cortex Underlie Distractibility and Suppression Deficits in Ageing," *Nature Communications*, Vol. 5, No. 4223, pp.1-27, June 2014.
- [4] B. G. Witmer, M. J. Singer, "Measuring Presence in Virtual Environments: A Presence Questionnaire" *Presence*, Vol. 7, No. 3, pp. 225-240, June 1998.
- [5] S. M. Slobounov, et al, "Modulation of cortical activity in 2D versus 3D virtual reality environments: an EEG study," *Int J Psychophysiol*, Vol. 95, No. 3, March 2015.
- [6] S. E. Kober, and C. Neuper, "Using auditory event-related EEG potentials to assess presence in virtual reality" *International Journal of Human-Computer Studies*, Vol. 70, No. 9, pp. 577-587, September 2012.
- [7] B. Evangelia, et al, "An EEG-based Evaluation for Comparing the Sense of Presence between Virtual and Physical Environments" *In Proc. Int Conf. Computer Graphics*. Bintan Island, Indonesia, Jun 11-14, 2018, pp. 107-116.
- [8] A. Crail, et al, "Deep learning for EEG classification tasks: A review," *Journal of Neural Engineering*, Vol. 16, pp. 1-38, April 2019.
- [9] V. J. Lawhern, et al, "EEGNet: A Compact Convolutional Neural Network for EEG-based Brain-Computer Interfaces," *Journal of Neural Engineering*, Vol. 15, No. 5, pp. 1-30, July 2018.
- [10] M. Ancona, et al, "Towards better understanding of gradient-based attribution methods for deep neural networks," *In Proc. 6th International Conference on Learning Representations*. Vancouver, BC, Canada, April 30-May 3, 2018, pp. 1-16.
- [11] C. E. Rolfe, et al, "Enhancing spatial attention and working memory in younger and older adults," *J Cogn Neuroscience*, Vol. 29, No. 9, pp. 1483-1497, Sep 2017.
- [12] M. I. Posner, "Orienting of attention" *Q J Exp Psychol*, Vol 32, No. 1, pp. 3-25, February 1980.
- [13] J. Polich, "Updating P300: An Integrative Theory of P3a and P3b," *Clin Neurophysiol*, Vol. 118, No. 10, pp. 2128-2148, Oct 2007.
- [14] J. D. Kropotov, "Chapter Executive System: A P3b component as index of engagement operation," in *Quantitative EEG, Event-Related Potentials and Neurotherapy*, 1st ed. San Diego, CA, USA: Academic, 2009, pp. 399-410.
- [15] H. C. Shan, Y. Liu, T. Stefanov, "A Simple Convolutional Neural Network for Accurate P300 Detection and Character Spelling in Brain Computer Interface," *In Proc. 27th International Joint Conference on Artificial Intelligence*. Ålvsjö, Sweden, July 13-19, 2018, pp. 1604-1610.